Punishment, Inequality and Emotions

David Masclet, Marie Claire Villeval

DISCUSSION PAPER SERIES

IZA DP No. 2119

Punishment, Inequality and Emotions

David Masclet
Marie-Claire Villeval

May 2006

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Punishment, Inequality and Emotions

David Masclet
CNRS, CREM, University of Rennes 1

Marie-Claire Villeval
GATE (CNRS, University of Lyon 2, ENS-LSH)
and IZA Bonn

Discussion Paper No. 2119
May 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

IZA Discussion Paper No. 2119
May 2006

ABSTRACT

Punishment, Inequality and Emotions*

Cooperation among people who are not related to each other is sustained by the availability of punishment devices which help enforce social norms (Fehr and Gächter, 2002). However, the rationale for costly punishment remains unclear. This paper reports the results of an experiment investigating inequality aversion and negative emotions as possible determinants of punishment. We compare two treatments of a public good game, one in which costly punishment reduces the immediate payoff inequality between the punisher and the target, and one in which it does not affect inequality. We show that while inequality-aversion prevents some subjects from punishing in the equal cost treatment, negative emotions are the primary motive for punishment. Results also indicate that the intensity of punishment increases with the level of inequality, and reduces earnings inequality over time.

JEL Classification: A13, C92, D63

Keywords: inequality aversion, negative emotions, free-riding, cooperation, experiment

Corresponding author:

Marie-Claire Villeval
GATE
93, Chemin des Mouilles
69130 Ecully
France
Email: villeval@gate.cnrs.fr

* This paper has benefited from useful comments from participants at the 2005 World Conference of the Society of Labor Economists and European Association of Labour Economists in San Francisco, the ESA Meeting in Montreal, and the 2nd International Meeting on Experimental and Behavioral Economics in Valencia. We are also grateful to R. Zeiliger for programming the experiment presented in this paper, and to the MiRe – DREES (French Ministry of Social Affairs) for a grant to support this research.

I. INTRODUCTION

Individuals sometimes cooperate with the authorities in reporting criminal activity, and witnesses accept to testify in favor of or against an unknown person even though they cannot expect any future benefits from doing so. The availability of costly sanctions has been shown to help enforce a social norm of cooperation among unrelated individuals (Fehr and Gächter, 2000 and 2002). [i] [ii] However, the forces driving subjects to make sacrifices in order to punish others, especially when they do not expect to become involved in long-run interactions, remain unclear.

This paper contributes to the literature on the behavioral determinants of altruistic punishment by examining the relative importance of inequality-aversion as an explanation of punishment in the framework of a public good experiment. The desire to induce higher contributions to the public good, and thus ensure a higher individual return in future periods, constitutes a strategic motive for punishment. However, since subjects also punish when they are certain that they will not interact again with the same person in the future, non-strategic motives, such as emotions and distributional concerns, have been suggested as alternative explanations.

In a cooperation game, negative emotions such as anger or disapproval are one non-strategic motive for individuals to sacrifice in order to punish others, even when there are no strategic reputation gains from doing so. Cooperators punish those who violate the pro-social norm of cooperation, or defectors punish those who try to establish such a norm (Fehr and Gächter 2000; Gintis 2000; Masclet, Noussair, Tucker and Villeval 2003; Carpenter and Matthews 2004; Carpenter, Matthews and Ong'ong'a 2004; and Gächter and Hermann 2005). Anderson and Putterman (2006) show that while punishing behavior follows the law of demand, subjects will punish even if there is a higher cost to the punisher than to the target. While Fehr and Gächter

(2002) show that even selfish subjects have reasons to cooperate because they anticipate punishment from altruistic punishers, Hopfensitz and Reuben (2005) emphasise the importance of social emotions such as shame and guilt in the punished as an essential component for the successful enforcement of cooperation. Such emotion-based punishment underscores the importance of intentions (Rabin 1993; Dufwenberg and Kirchsteiger 2004). The importance of emotions in decision-making has also been established by neuroscientists (Damasio 1994) and research in neuroeconomics has started to provide some evidence of a neural basis for altruistic punishment (Quervain, Fischbacher, Treyer, Schellhammer, Schnyder, Buck and Fehr 2004).

A second non-strategic reason to punish group members is inequality aversion (Fehr and Schmidt 1999; Falk, Fehr and Fischbacher 2006; Bolton and Ockenfels, 1999).[iii] In economics, the relationship between punishment and inequality aversion has been far less scrutinized than that with emotions.[iv] In contrast, the fitness differential theory in evolutionary psychology posits that punishment aims to reduce the payoff advantage of free-riders (Price, Cosmides and Tooby 2002). Individuals with distributional concerns who suffer from disadvantageous inequality may be willing to pay to punish defectors in order to reduce earnings inequality, if the cost they bear themselves is smaller than the impact of sanctions on the target's payoff. This suggests that we should take a closer look at the influence of the feelings caused by inequality on punishment.

The contribution of our paper is threefold. We first examine the relative importance of inequality aversion and emotions among the non-strategic motives for punishment in a public good game. If motivated by emotions, an individual may be willing to undertake costly punishment even when punishment does not affect the distribution of earnings. In contrast, if only motivated by inequality aversion, individuals will not punish when punishment does not change the distribution

of earnings. Our experiment consists of two treatments of a public good game, implemented with a stranger matching protocol. In the Unequal Cost Treatment, punishment affects payoff inequality between the punisher and the target (replicating the design of Fehr and Gächter, 2000), and in the Equal Cost Treatment it does not affect the current-period level of inequality. In the second treatment the ratio between the cost of one punishment point to the punisher and the cost of this point to the target equals one. On the contrary, in the Unequal Cost Treatment, this ratio should be greater than one in most circumstances.[v] The strong prediction is that subjects who only care about their own payoff and earnings inequality should then not punish in the Equal Cost Treatment. This prediction is based on Fehr and Schmidt's (1999) model. A weaker hypothesis allows several different motives to play a role; the comparison between treatments allows us to identify the importance of the concern for inequality in punishment behavior.

Second, we measure the importance of strategic motives in punishment by repeating the Equal Cost Treatment in a partner-matching protocol in which subjects repeatedly play the public good game with the same group. Comparing behavior across the two matching protocols allows us to separate strategic from non-strategic punishment behavior, and to differentiate between related and unrelated individuals. While any immediate effect of punishment on inequality is ruled out in this treatment, the design does help us to see whether the demand for punishment is partly explained by forward-looking distributional considerations (i.e. the desire to reduce the payoff advantage accruing to free-riders in the future).

The third contribution of this paper is to examine the indirect relationships between inequality and punishment. We also investigate the impact of punishment on the evolution of inequality.

Will punishment reduce the level of inequality over time by inducing free-riders to increase their contributions after being punished?

In contrast with most papers on punishment, we focus on the distance between the punisher's and the target's payoffs, rather than on the distance between the target's payoff and the average group payoff. We are close to Falk, Fehr and Fischbacher (2006) who analyze the importance of fairness and spite in punishment in three-player one-shot Prisoner's Dilemma games. Spiteful punishment is carried out by subjects who value the target's payoff negatively, irrespective of the distribution of pre-punishment payoffs. They conclude that punishment is not primarily driven by the willingness to change payoff shares, but by the desire to retaliate and harm those who behave unfairly. Here we consider another type of cooperation game, a four-person public-good experiment, in which the set of actions is larger.

Our first finding is that our experiment rejects the strong assumption that punishment is only motivated by inequality aversion, but accepts the weaker assumption of punishers paying some attention to the distribution of payoffs. Subjects do punish even in the Equal Cost Treatment. However, the comparison across treatments indicates that a significant number of subjects refrain from punishing when they cannot affect earnings inequality by doing so. The second finding shows the importance of negative emotions. If we control for selection bias, as some subjects do not punish, and the level of inequality between the punisher and the punished, we find that cooperators punish more heavily in the Equal than in the Unequal Cost Treatment. This indicates that they are willing to impose the same payoff reduction as in the Unequal Cost Treatment, since the monetary impact of each punishment point on the target is smaller than in the Unequal Cost Treatment. This finding is consistent with negative emotions: feelings of anger provoke greater

expenditure in punishment since the punishers focus more on the consequences of the sanction on the target than on their personal cost of punishing.

A third finding is that although subjects are primarily motivated by negative emotions, these emotions arouse from the observation of inter-individual inequality in contributions and earnings. Fourth, people do not punish more in the Equal Cost Treatment played under the partner-matching protocol. We observe however that the efficiency of punishment is greater in fixed groups, underscoring the importance of strategic motives with repeated interactions. Last, punishment reduces inequality over time, by inciting free-riders to increase their effort in following periods, which reduces the future inequality in payoffs to cooperators. We therefore cannot reject that in repeated interactions subjects punish in order to reduce inequality over time.

The remainder of the paper is organized as follows. Section 2 details our experimental design and the procedures, and Section 3 presents the theoretical predictions of the model, with either purely selfish agents or in the presence of agents with distributional concerns. Section 4 displays the results and Section 5 concludes.

II. EXPERIMENTAL DESIGN AND PROCEDURES

Design. The experiment is based on a public good game, and involves groups of four subjects. It consists of 30 periods, divided into three segments. In the first ten periods and the last ten periods, subjects play a standard public good game. This no-punishment condition serves as a benchmark for the condition with punishment opportunities which occurs from periods 11 to 20. At the beginning of each period, each group member is endowed with 20 units. Each member simultaneously selects a fraction of her endowment to contribute to a group account, while

keeping the remainder in her private account. All funds in the group account pay a positive return to each member. The parameters are chosen so that full free-riding is a dominant strategy whereas full contribution to the public good corresponds to the social optimum.

First consider the condition without punishment. In periods 1-10 and 21-30, each subject *i* chooses a fraction g_i of his endowment as a contribution to the public good. All group members' decisions regarding g_i are made simultaneously. The marginal per capita return from a contribution to the group account is 0.4. Subject *i* 's payoff is given by:

$$\pi_i = 20 - g_i + .4\sum_{j=1}^{4} g_j \qquad (1)$$

The group members are informed of both the amount of the group contribution and their individual payoff.

Now consider the condition with punishment. Each period 11-20 of both treatments consists of a two-stage game. The first stage is identical to that in the previous condition: each group member receives an endowment and has to decide how much to invest in the group account. In the second stage, each subject, after being informed about each other group member's contribution, can assign 0 to 10 punishment points to any of the other three group members. Assigning points is costly. The main difference between the Unequal and Equal Cost Treatments lies precisely in the monetary consequences of punishment points. The schedule of costs in these two treatments is given in Table 1.

In the Unequal Cost Treatment, each punishment point received from any other subject reduces first-stage earnings by 10%, up to a maximum of 100%. This treatment intends to replicate that in Fehr and Gächter (2000). As usual, contributions are listed in random order and with a different

identification number on the screen each period so that it is impossible to target another subject for punishment for more than one period. This rules out motivations such as revenge. Subject *i*'s earnings per period are now given by:

$$\left(20-g_i+0.4*\sum_{k=1}^{n}g_k\right)*\frac{\max\left\{0,10-\sum_{k\neq i}P_{ki}\right\}}{10}-\sum_{k\neq i}C(P_{ik}) \qquad (2)$$

where P_{ik} is the number of points assigned by i to k, and $C(P_{ik})$ is the cost to *i* of assigning the points to *k*. The design of punishment is such that if a subject who plays the Nash equilibrium is punished, the payoff inequality is necessarily reduced between the punisher and the target. In most circumstances it guarantees a reduction of payoff inequality.[vi] Losses are possible if the cost of punishing others exceeds the individual's net income but they are extremely unlikely (this situation never occurred in our experiment).

In the Equal Cost Treatment, each point given by a punisher has the same monetary cost for himself and the target: the cost of being punished always equals the cost of punishing. Suppose that subject *i* assigns say 3 punishment points to subject *j*; the first-stage earnings of both subject *i* and subject *j* are reduced by 4 units, as described in Table 1. Subject *i*'s earnings per period in this treatment are given by:

$$\left(20-g_i+0.4*\sum_{k=1}^{n}g_k\right)-\sum_{k\neq i}C\left(P_{ki}\right)-\sum_{k\neq i}C\left(P_{ik}\right) \qquad (3)$$

Procedures. The experiment was computerized using REGATE software (Zeiliger, 2000). We ran eight sessions in total under the stranger matching protocol, split 50:50 between the Equal and Unequal Cost Treatments. In total 72 subjects participated in each treatment. We also ran two sessions with the partner matching protocol under the Equal Cost Treatment, with 12

participants each. Six sessions were conducted in the experimental laboratory of the Groupe d'Analyse et de Theorie Economique (GATE) at the University of Lyon, and four sessions were organized in the LABoratory of EXperimental Economics at the University of Rennes, France. In total, 168 subjects were recruited from undergraduate classes in business and engineering schools in Lyon and in various departments in Rennes. We did not recruit any economics students, and none of the subjects had any experience of this particular type of experiment. No subject participated in more than one session.

Upon arrival, the subjects drew a label from a bag, indicating the name of their computer. The instructions (see the Appendix) were distributed and read aloud. The subjects then filled out a questionnaire that allowed us to check their understanding of the rules of the game. Questions were answered in private. The program then matched subjects randomly and anonymously. Under the stranger matching protocol, groups were reshuffled after each period, whereas under the partner matching protocol, the composition of groups remained unchanged over time. During each ten-period segment, subjects did not know if the experiment would extend beyond the current segment.

On average sessions lasted for 90 minutes, including reading instructions and payment. Each unit was convertible to Euro at 100 units = 2 Euros. Each participant received €16.40 on average, including a show-up fee of €3.

III. THEORETICAL PREDICTIONS

If players are selfish, the unique subgame-perfect equilibrium of both the Unequal and Equal Cost Treatments is to contribute nothing in each period and never to punish. Punishment is not

credible in either treatment. Therefore, complete free-riding is a dominant strategy in all periods and all treatments.

Relaxing the selfishness assumption, consider the predictions of Fehr and Schmidt's (1999) inequality-aversion theory. Here individual utility depends not only on ones own payoffs but also on the equality of the income distribution. Individuals are inequality averse if they incur disutility both from being worse off in material terms than others (disadvantageous inequality) and from being better off than others (advantageous inequality). This disutility is self-centered in the sense that subject *i* compares herself to each of the other group members but does not care about inequalities between the other members. Moreover subjects are assumed to be more sensitive to disadvantageous inequality, as shown by the inequality-aversion term α_i, $\alpha_i > 0$, than advantageous inequality, given by β_i with $0 \leq \beta_i < 1$ such that $\alpha_i > \beta_i$. The utility function of player $i \in \{1,\ldots,n\}$ is:

$$U(x_i) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max(x_j - x_i, 0) - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max(x_i - x_j, 0) \qquad (4)$$

The first term of equation (4) represents the pecuniary payoff of subject *i*. The second and the third terms measure the utility loss from disadvantageous and advantageous inequality respectively. Following Fehr and Schmidt, we assume that the utility function is linear in both inequality aversion and x_i. This implies that the marginal rate of substitution between monetary income and inequality is constant.

To solve the standard public good game, we first substitute equation (1) (the monetary payoffs x_i and x_j) into the utility function (4). Without any chance of punishment, if subjects are sufficiently averse to disadvantageous inequality, and if advantageous inequality aversion is sufficiently low

relative to marginal earnings (i.e. if $1-\alpha > \beta_i$), then not contributing is a Nash equilibrium. When subjects cannot punish each other, cooperation cannot be enforced, even if they are inequality averse. The only way for inequality-averse subjects to reduce such inequality is by reducing their own contributions (the formal proof is provided in Fehr and Schmidt, 1999).

Now consider punishment in the Unequal Cost Treatment. We substitute the monetary payoffs of *i* and *j* from equation (2) into equation (4). In the second stage, punishment strategies are credible since no enforcer gains by not punishing. If he does not punish, he reduces his expenditure but suffers from both disadvantageous inequality relative to the defector and advantageous inequality relative to the enforcers. Under sufficient conditions, subjects will always choose to punish. Anticipating credible punishment in the second stage, no-one free rides in the first stage. Anyone deviating by free riding in stage 1 will be punished by the other group members in stage 2. Here, disutility from inequality can be reduced since players can affect the payoff of others by punishing them (see the formal proof in Fehr and Schmidt, 1999).

Do these predictions hold when the cost of punishment is the same for the punisher and the target, as in the Equal Cost Treatment? Inequality-aversion predicts no punishment and no cooperation in this treatment. If a cooperator punishes a defector, whereas the other subjects play the equilibrium in the second stage, she incurs the cost of the punishment but does not change the earnings gap between herself and the target. In addition, she suffers from disadvantageous inequality relative to the cooperators who do not punish.

Last consider the case of subjects who are neither selfish nor inequality averse, but driven by emotions. They may punish a norm violator even though punishment will not affect the level of inequality between themselves and the target. With a sufficient number of subjects driven by

anger or disapproval, we may observe no difference in punishment between the Equal and Unequal Cost Treatments.

V. EXPERIMENTAL RESULTS

We first examine the impact of the various punishments on contribution decisions. We then analyze punishment behavior to identify inequality considerations and negative emotions, before studying strategic punishment. We then focus on inter-personal comparisons as a source of norms, and last show the effect of punishment on the evolution of inequality over time.

Contribution Decisions

We look at contributions to the public good by period and by treatment to see whether behavior changes when punishment is possible and, if so, whether it is influenced by the treatment and the matching protocol. Figure 1 displays the evolution of average contributions over time in both the Unequal and Equal Cost Treatments, under both the stranger and partner matching protocols.

In all of the treatments the data exhibit the same pattern as in Fehr and Gächter (2000): in periods 1-10, subjects initially contribute more than the Nash equilibrium level, but progressively reduce their contributions; in periods 11-20 the introduction of the punishment opportunity entails an increase in average contributions; last, in periods 21-30, average contributions drop off sharply when punishment opportunities are withdrawn.

Result 1: *In all treatments, punishment causes an increase in average contributions.*

Support for result 1: In the Equal Cost Treatment under the partner matching protocol, a Wilcoxon matched pair test rejects the null hypothesis that contributions are identical between periods 11-20 (punishment) and the pooled data from periods 1-10 and 21-30 (no punishment)

(p=.027).[vii] In the Unequal cost treatment under the stranger protocol, an analogous Wilcoxon matched pair test also rejects this null hypothesis (p=0.0679). The same test for the Equal cost treatment with stranger matching also reveals a significant difference between periods 11-20 and the pooled data from periods 1-10 and 21-30 ($p<0.1$).

In the stranger matching protocol, a Mann-Whitney rank sum test accepts the null hypothesis that there is no difference in contributions in periods 11-20 between the Unequal and Equal Cost Treatments (p=.309). The same test comparing contributions in the stranger and partner conditions for the Equal Cost Treatment reveals that subjects contribute significantly more under repeated interactions ($p<0.05$).

Punishment behavior

We next examine the distribution of punishment points by subjects to other group members to determine whether behavior differs according to the treatment and the matching protocol. Table 2 shows the evolution of the average number of points distributed over time. In all treatments, subjects use costly punishment, even in the one-shot game, but the average number of punishment points decreases over time. This is consistent with results from previous work. We also show that subjects punish even in the Equal Cost Treatment.

Result 2: *Subjects punish even when they cannot alter the distribution of payoffs.*

Support for result 2. In any period of the stranger condition, 36.4% of subjects distribute at least one punishment point in the Unequal Cost Treatment; the analogous figure in the Equal Cost Treatment is 32.2%, and in the partner condition of the Equal Cost Treatment it is 37.1%. In the stranger condition, a Mann-Whitney rank sum test accepts the null hypothesis that there is no

difference in the proportion of punishers between the Unequal and Equal Cost Treatments (p=.248). The same conclusion results from the comparison between the stranger and partner conditions in the Equal Cost Treatment (p=.199).

These statistics confirm that subjects use punishment even when it has no effect on inequality. However, they do not take into account the amount contributed in groups. Further, it is important to dissociate the decision to punish from the intensity of punishment: if partly motivated by inequality-aversion, some subjects may abstain from punishing, which introduces a potential selection bias in the estimation of the number of punishment points.

Result 2': *Conditional on the willingness to punish, subjects who punish defectors do so more intensely in the Equal than in the Unequal Cost Treatment.*

Support for result 2': To analyze the sensitivity of the number of punishment points distributed in periods 11 to 20 to the experimental conditions, we estimate a model including correction for selection bias. As the treatment may have different effects on the decision to punish and the intensity of punishment, we use a two-step estimation procedure. We consider punishment probability using a random-effects Probit model; we then explain the number of points, conditional on the punishment decision, by a Generalized Least Squares model corrected for selection bias via the inverse of the Mills ratio (the "IMR" variable). Each regression measures the influence of the Equal Cost relative to the Unequal Cost Treatment in the stranger matching protocol. The exogenous variables in the selection equation include the positive and negative deviations between the target's contribution and both the punisher's and average group contributions[viii], as well as a period variable. The GLS regression includes the same variables except for the period variable, which allows us to identify the model.

The estimation results are shown in Table 3. Column (1) presents the selection results and column (2) the marginal effects. Column (3) shows the results of the GLS estimation for the whole population. The last two columns give the results of the GLS estimations for the sub-samples of observations in which the subject punishes a group member who contributes less than himself (column (4)) or more than himself (column (5)).

Controlling for selection bias and for relative contribution levels, the regressions show that, first, subjects are significantly less likely to punish in the Equal than in the Unequal Cost Treatment and, second, that those who decide to punish do so more intensely in the Equal Cost treatment.

This shows that inequality aversion does play some role, since some subjects do not punish when payoff shares cannot be altered (see regression (1)). However, the marginal effect of the treatment is small: the Equal Cost Treatment reduces the probability of punishment by 4% (see column (2)).

We can also see the strength of negative emotions. The regression in column (3) shows that, conditional on the willingness to punish, subjects are willing to pay more to increase the harm imposed on targets in the Equal Cost Treatment. This is because the monetary consequence of each punishment point on the target is lower in the Equal than in the Unequal Cost Treatment. Indeed, on average the monetary consequence of a punishment point is 2.7 times lower in the Equal than in the Unequal Cost Treatment. This is consistent with the results obtained by Falk, Fehr and Fischbacher (2006) in PD games where cooperators increase their punishment expenditures if the impact of punishment on the targets is reduced. Therefore, while Anderson and Putterman (2006) have shown that punishment behavior obeys the law of demand,

individuals take into account not only the cost to themselves but also the cost imposed on the target. They are ready to pay more if the "quality" of punishment is higher. Negative emotions lead them to pay a higher price in the Equal Cost Treatment to increase the consequences of their sanction on the targets. Interestingly, this is only true for the subjects who punish those who contribute less than themselves (regression (4)). In contrast, the subjects who punish those who cooperate more and are not inequality averse, by definition, are not willing to pay more to punish in the Equal than in the Unequal Cost Treatment (regression (5)). An interpretation is that the intensity of their negative emotions is lower than that of the cooperators *vis-à-vis* free-riders. Again this result is consistent with that observed in PD games.

We now consider the strategic motive for punishment. To isolate this motive, we measure the impact of the protocol on the decision to distribute punishment points which do not affect the distribution of payoffs. We again use a two-step procedure to account for potential selection bias. In Table 4, column (1) presents the results of a random-effects Probit model analyzing the probability of punishment and column (2) shows the marginal effects. Column (3) shows the GLS estimation of the number of punishment points corrected for selection. Column (4) shows the same model estimated on the sub-sample of observations in which a subject punishes those who contribute less than himself.[ix]

In Table 4 the nature of the interaction has no significant effect on either the decision to punish or the intensity of punishment. This suggests that pressure is exerted as strongly between unrelated individuals (the stranger condition) as between related individuals (the partner condition).

Result 3: *The strategic motive for punishment is less important than non-strategic motives.*

As in PD games (Falk, Fehr and Fischbacher 2006), people punish mostly to harm targets in retaliation for the negative emotions they have aroused. The reduction of payoff differentials is a secondary motive and when sanctions cannot affect the payoff shares there is no difference between the one-shot and the repeated game.

Inequality and the arousal of emotions

If people allocate punishment points even when they do not affect the earnings gap with the punished, this does not necessarily mean that they do not care about the level of inequality. On the contrary, the observation of inter-individual differences in contributions brings about negative emotions.

Result 4: *Inter-individual comparisons cause negative emotions and punishment*.

Support for result 4. Figure 2 displays the distribution of punishment points as a function of the level of inequality between the punisher and the target, by treatment and protocol. This difference represents the absolute level of inequality in both contributions and earnings at the end of the first stage of each period.[x] The relative importance of each category of deviation is indicated above each bar.

This figure clearly shows that punishers react strongly to the observed level of inequality under all conditions. Subjects who contribute less and earn more are punished much heavily than those who earn almost the same payoff after the first stage. For example, a subject who earns between 14 and 20 units more than the punisher receives on average 1.9 points of punishment from this punisher whereas he receives only .07 point if he contributes and earns almost the same amount as the punisher (within the [-2,2] interval).

This is formalised in the econometric analysis. The regressions reported in Table 3 show that the distribution of punishment points is less likely when an individual contributes more and earns less than the potential punisher. In contrast, disadvantageous inequality between the punisher and the target increases both the probability of punishment and the number of punishment points distributed. The regressions reported in Table 4 confirm that the intensity of punishment is influenced by inter-individual deviations in contributions and earnings; they also show that individuals are not punished for deviations from average group behavior. This confirms that inter-individual inequality gives rise to negative emotions and retaliation. Inter-individual comparisons are more relevant than comparisons between the individual and average behavior and earnings within the group.

We also see some subjects who punish cooperators who are already suffering from disadvantageous inter-personal inequality. In this sub-sample (see regression (5) in Table 3), the number of punishment points distributed increases in the deviation of the target's contribution from the average and not from the punisher's contribution. This suggests that free-riders' negative emotions are triggered by the fact that cooperators raise the social norm, and not by inter-individual comparisons. This is consistent with Gächter and Hermann (2005).

Overall, these results show that while negative emotions seem to be the main reason for punishment, rather than inequality reduction, inter-individual comparisons and inequality are an important source of the negative emotions that trigger the desire to punish.

Consequences of punishment on the level of inequality

How does punishment change the level of inequality in groups over time? We suspect that even if punishment does not directly affect inequality, punishment in one period may reduce inequality in further periods if it affects future contributions by the punished.

Result 5: *Under all conditions, individuals who contributed less than average in period t increase their contributions more in period t+1 the more they were punished in t. As a result of both the credible threat of punishment and its actual impact, payoff inequality decreases over time.*

Support for result 5: Table 5 displays the estimates of random-effects GLS models analyzing the determinants of the change in contribution between periods *t* and *t+1*. These are estimated separately for those who contributed less than average and more than average (on the left and right sides of the Table respectively). The first columns correspond to the pooled data from the Equal and Unequal Cost Treatments under the stranger matching protocol; the second columns correspond to the pooled data from the Equal Cost Treatment under both matching protocols; and the third columns pool all data together. The explanatory variables include the punishment received in period *t* and the deviation from the average contribution of other group members.[xi] We also interact the punishment points received with the treatment and with the protocol. These interactions show whether punishment is more or less effective under the Equal Cost treatment, and under the partner matching protocol.

Table 5 reveals a significant negative relationship between deviation from the average contribution and the subsequent change in contributions. In contrast, punishment boosts the contributions of those who contributed less than average in all treatments. Those who contributed more than average are unaffected. The Table also shows that a punishment point is

more effective under the Unequal than under the Equal Cost Treatment. As we saw earlier, this can be explained by the smaller monetary consequences of each punishment point on the target in the Equal Cost Treatment, which is anticipated by punishers who punish a given deviation more heavily. Finally, the regressions show that subjects react more to punishment in the partner-matching protocol, underlining the importance of strategic motives for contributing in repeated interactions.

How do these results translate into payoffs and inequality over time? Table 6 displays the evolution of average payoffs and Gini coefficients over time with and without punishment. To avoid possible biases due to restart and end-game effects, we do not compare directly the levels of inequality in the first and in the last period. Instead, we consider pooled data over three periods at the beginning and at the end of each set of rounds played under the same treatment.

At the aggregate level, Wilcoxon tests do not reject the null hypothesis that the Gini index is similar for treatments with and without punishment played under a stranger matching protocol. Considering disaggregated data by sets of periods, Table 6 shows that inequality declines over time within each set of periods for each treatment.

We also observe that when punishment is not possible, free-riding leads to more equal payoffs but with falling earnings. On the contrary, punishment produces falling inequality along with a slight tendency to rising earnings (except in the Equal Cost Treatment under the stranger matching protocol where earnings are flat).[xii] For example, in the Unequal Cost Treatment without punishment, the average payoff drops from 23.15 to 20.8, whereas with punishment average earnings rise from 20.70 at the beginning of the set of periods to 23.06 at the end of the periods.

V. DISCUSSION AND CONCLUSION

Previous research has shown that punishment is an important mechanism underlying cooperation among unrelated individuals, but the reasons why individuals are willing to pay to punish others remain unclear. In this paper we have looked at the influence of inequality aversion on the punishment decision, compared to that of negative emotions and strategic motives. We have considered a number of treatments in a public good game in which we introduce punishment opportunities with different consequences on the distribution of payoffs. Our econometric analysis allows us to separate the decision to punish from the intensity of punishment, so that we can control for possible selection bias.

We reject the strong hypothesis that individuals only punish to reduce inequality between themselves and the targets, but support the weak hypothesis of some influence of inequality aversion on punishment. However, the marginal effect of the nature of the punishment on the decision to punish shows that punishment results more from negative emotions than from inequality aversion or strategic motives (i.e. increasing the target's contribution in future periods). Introducing punishment that does not affect the distribution of payoff shares reduces the probability of punishment by only 4%. These results are remarkably consistent with those obtained in one-shot PD games by Falk, Fehr and Fischbacher (2006).

While punishers are not primarily motivated by inequality-aversion, nonetheless, in all treatments, inter-individual comparisons are used as a social norm to decide whom to punish. Negative emotions are primarily triggered by the observation of inequality in contributions and earnings more than by the lack of adherence to the social norm of contributing. Thus, instead of contrasting inequality aversion and negative emotions, further research is needed to understand

how inequality may give rise to negative emotions. Another implication is that we should concentrate more on inter-individual comparisons than on comparisons between individual and average behavior in groups.

Finally, we also show that punishment within a group reduces inequality over time. This suggests that, when groups are fixed, team members might punish strategically not only to increase their own payoff in future periods, but also to reduce earnings inequality over time, even though punishment leaves current-period inequality unaffected. This impact of punishment on inequality over time is also seen in groups of unrelated individuals. We find evidence that free-riding progressively gives rise to reduced inequality coupled with low earnings; in contrast, the disciplinary effect of punishment brings about falling inequality associated with higher earnings.

REFERENCES

Anderson, Christopher M. and Putterman, Louis, (2006). "Do Non-Strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism." *Games and Economic Behavior*, *54*(1), 1-24.

Bochet, Olivier; Page, Talbot and Putterman, Louis, (Forthcoming). "Communication and Punishment in Voluntary Contribution Experiments." *Journal of Economic Behavior and Organization*.

Bolton, Gary E. and Ockenfels, Axel, (2000). "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, *90*(1), 166-93.

Carpenter, J.P., (2006). "Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods." *Games and Economic Behavior*, (Forthcoming).

Carpenter, J.P. and Matthews, P., (2004). "Social Reciprocity." *Working Paper*. Middlebury College.

Carpenter, J.P.; Matthews, P. and Ong'ong'a, O., (2004). "Why Punish? Social Reciprocity and the Enforcement of Pro-Social Norms." *Journal of Evolutionnary Economics*, *14*.

Damasio, Antonio, (1994). *Descartes' Error - Emotion, Reason and the Human Brain*. New-York: Grosset - Putnam.

Dufwenberg, Martin and Kirchsteiger, Georg, (2004). "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, *47*, 268-98.

Engelman, Dirk and Strobel, Martin, (2004). "Inequality Aversion, Efficiency and Maximin Preferences in Simple Distribution Experiments." *American Economic Review*, *94*(4), 857-68.

Falk, Armin; Fehr, Ernst and Fischbacher, Urs, (2006). "Driving Forces Behind Informal Sanctions." *Econometrica*, *Forthcoming*, 1-14.

Fehr, Ernst and Gächter, Simon, (2002). "Altruistic Punishment in Humans." *Nature*, *415*(10), 137-40.

____, (2000). "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, *90*(4), 980-94.

Fehr, Ernst and Rockenbach, Bettina, (2003). "Detrimental Effects of Sanctions on Human Altruism." *Nature*, *422*, 137-40.

Fehr, Ernst and Schmidt, Klaus M., (1999). "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics*, *114*, 817-68.

Gächter, Simon and Hermann, Benedikt, (2005). "Norms of Cooperation among Urban and Rural Dwellers: Experimental Evidence from Russia." *University of Nottingham, mimeo*.

Gintis, Herbert, (2000). "Strong Reciprocity and Human Sociality." *Journal of Theoretical Biology*, *206*, 169-79.

Hopfensitz, Astrid and Reuben, Ernesto, (2005). "The Importance of Emotions for the Effectiveness of Social Punishment." *Tinbergen Institute Discussion Paper*, TI 2005 - 075/1. Amsterdam.

Houser, Daniel; Xiao, Erte; McCabe, Kevin and Smith, Vernon, (2005). "When Punishment Fails: Research on Sanctions, Intentions and Non-Cooperation." *George Mason University Working Paper*.

Masclet, David; Noussair, Charles; Tucker, Steve and Villeval, Marie-Claire, (2003). "Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review*, *93*(1), 366-80.

Price, Michael E.; Cosmides, Leda and Tooby, John, (2002). "Punitive Sentiment as an Anti-Free Rider Psychological Device." *Evolution and Human Behavior*, *23*, 203-31.

Quervain, D.J.F.; Fischbacher, Urs; Treyer, V.; Schellhammer, M.; Schnyder, U.; Buck, A. and Fehr, E., (2004). "The Neural Basis of Altruistic Punishment." *Science*, *305*, 1254-58.

Rabin, Matthew, (1993). "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, *83*, 1281-1302.

Zeiliger, Romain, (2000). *A Presentation of Regate, Internet Based Software for Experimental Economics*. http://www.gate.cnrs.fr/~zeiliger/regate/RegateIntro.ppt., GATE.

Table 1. Levels of punishment and associated costs

Unequal Cost Treatment

Punishment Points	0	1	2	3	4	5	6	7	8	9	10
Cost to the punisher in units	0	1	2	4	6	9	12	16	20	25	30
Cost to the target in % of the target's earnings from the 1st stage	0	10	20	30	40	50	60	70	80	90	100

Equal Cost Treatment

Punishment Points	0	1	2	3	4	5	6	7	8	9	10
Cost to the punisher in units	0	1	2	4	6	9	12	16	20	25	30
Cost to the target in units	0	1	2	4	6	9	12	16	20	25	30

Table 2. Level of punishment over time

Periods	Unequal Cost (Stranger)	Equal Cost (Stranger)	Equal Cost (Partner)
11	1.50	1.40	1.96
12	1.31	1.33	1.08
13	0.85	1.06	0.71
14	0.85	0.75	0.67
15	0.68	0.67	0.29
16	0.50	0.69	1.21
17	0.54	0.53	0.96
18	0.47	0.68	0.71
19	0.46	0.76	0.58
20	0.38	0.79	0.58

Table 3. Punishment (probability and intensity) in the Stranger Matching protocol

Variables	Random effects Probit model		GLS models		
	(1)	(2)	(3)	(4)	(5)
Equal Cost Treatment	-.416*** (.140)	-.036***	.488*** (.121)	.493*** (.134)	-.099 (.199)
Negative difference from the average	.022 (.023)	.002	-.021 (.187)	-.023 (.045)	.007 (.038)
Positive difference from the average	-.227*** (.020)	-.019***	-.037* (.023)	-.040* (.023)	.843*** (.154)
Average group contribution	-.032** (.013)	-.003**	-.015 (.013)	-.029** (.014)	-.003 (.034)
Negative difference from the punisher	.361*** (.016)	.031***	.112*** (.029)	.101*** (.030)	
Positive difference from the punisher	-.042** (.018)	-.004**	.085*** (.024)		.046 (.032)
Period	-.114*** (.013)	-.010***			
Constant	.180 (.276)		1.224*** (.310)	1.519*** (.327)	.205 (.897)
ρ	.496 (.040)				
IMR			-.210* (.127)	-.308** (.132)	.322 (.415)
Nb observations	4320		668	597	59
Log Likelihood	-1119.615				
R^2			.258	.261	.551
Wald χ^2	711.29		238.77	242.54	64.17
p> χ^2	.0000		.0000	.0000	.0000

Note: standard errors in parentheses. *** statistically significant at the .01 level; ** at the .05 level; * at the .10 level. Each individual appears 30 times (1 observation for each pair within the group over 10 periods)

Table 4. Punishment (probability and intensity) in the Equal Cost Treatment (pooled data from both protocols)

Variables	Random effects Probit model		GLS models	
	(1)	(2)	(3)	(4)
Partner matching protocol	.072 (.239)	.005	.079 (.206)	.162 (.220)
Negative difference from the average	.002 (.032)	.0001	-.014 (.044)	-.0202 (.059)
Positive difference from the average	-.233*** (.024)	-.016*** 	-.005 (.039)	-.007 (.039)
Average group contribution	-.003 (.015)	-.0002	-.029* (.018)	-.042** (.019)
Negative difference from the punisher	.318*** (.018)	.022*** 	.112** (.046)	.100** (.046)
Positive difference from the punisher	-.140*** (.032)	-.009*** 	.113** (.058)	
Period	-.094*** (.016)	-.006***		
Constant	-.3821 (.338)		1.794*** (.519)	2.091*** (.534)
ρ	.430 (.050)			
IMR			-.29793 (.221)	-.400* (.227)
Nb observations	2880		410	384
Log Likelihood	-707.945			
R^2			.254	.252
Wald χ^2	461.93		161.12	161.70
p> χ^2	.000		.000	.000

Note: standard errors in parentheses. *** statistically significant at the .01 level; ** at the .05 level; * at the .10 level.

Table 5. The change in contributions between periods *t* and *t+1*

Dependent variable: Change in contribution between *t* and *t+1*	Targets who contributed less than the average			Targets who contributed more than the average		
	Stranger	Equal	Pooled data	Stranger	Equal	Pooled data
Points received in period *t*	.894*** (.126)	.506*** (.104)	.833*** (.125)	.653* (.388)	.220 (.523)	.631 (.414)
Points received * Equal Cost Treatm.	-.412*** (.111)		-.376*** (.116)	-.448 (.597)		-.441 (.639)
Points received * partner protocol		.251* (.146)	.238* (.130)		1.660 (1.823)	1.630 (1.742)
Deviation from the average	-.314*** (.060)	-.303*** (.077)	-.349*** (.055)	-.770*** (.054)	-.690*** (.077)	-.713*** (.054)
Constant	-.522** (.215)	-.398 (.317)	-.492** (.211)	.277 (.288)	-.214 (.376)	-.002 (.280)
Observations	616	380	692	634	434	744
R^2	.308	.276	.322	.207	.150	.166
Wald χ^2	287.68	157.63	352.27	203.82	82.49	174.69
$p > \chi^2$	.000	.000	.000	.000	.000	.000

Note: standard errors in parentheses. *** statistically significant at the .01 level; ** at the .05 level: * at the .10 level.

Table 6. Changes over time in average payoffs and Gini coefficients by treatment

Groups of periods	All periods		Periods without punishment		Periods with punishment	
	Without punishment	With punishment	1-3 and 21-23	8-10 and 28-30	11-13	18-20
Unequal Cost Strangers						
Mean Payoff	21.81	22.51	23.15	20.80	20.70	23.06
Gini	.10	.09	.12	.06	.13	.08
Equal Cost Strangers						
Mean Payoff	22.09	21.97	23.37	21.17	21.94	20.93
Gini	.10	.10	.11	.08	.13	.10
Equal Cost Partners						
Mean Payoff	22.22	25.66	24.02	20.67	24.08	25.72
Gini	.10	.08	.11	.06	.09	.08

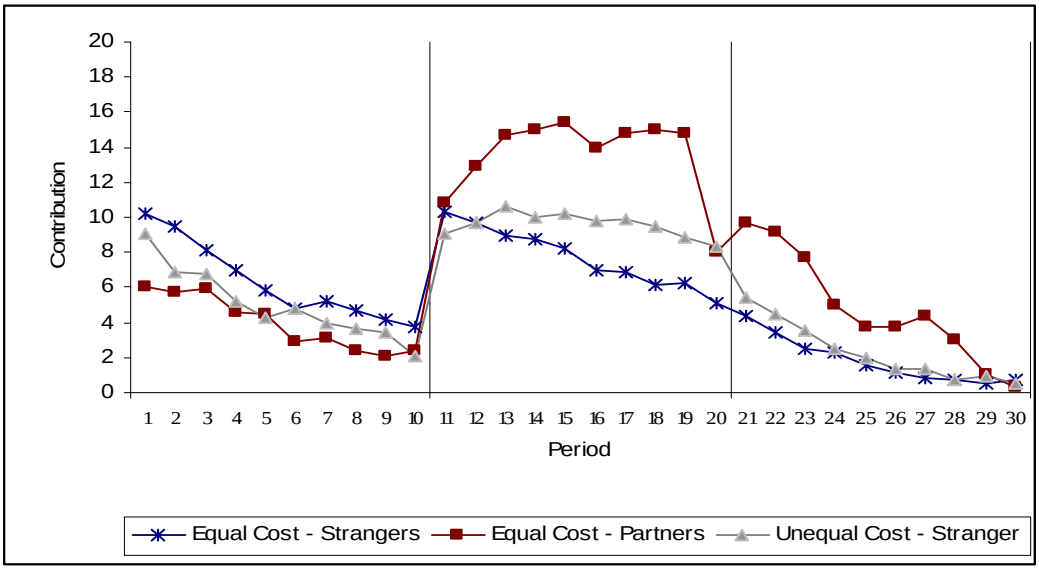

Fig.1. Average contributions over time

4% 1% 2% 4% 6% 5% 11% 24% 23% 62% 38% 40% 11%24% 23% 4% 6% 5% 4% 1% 2%

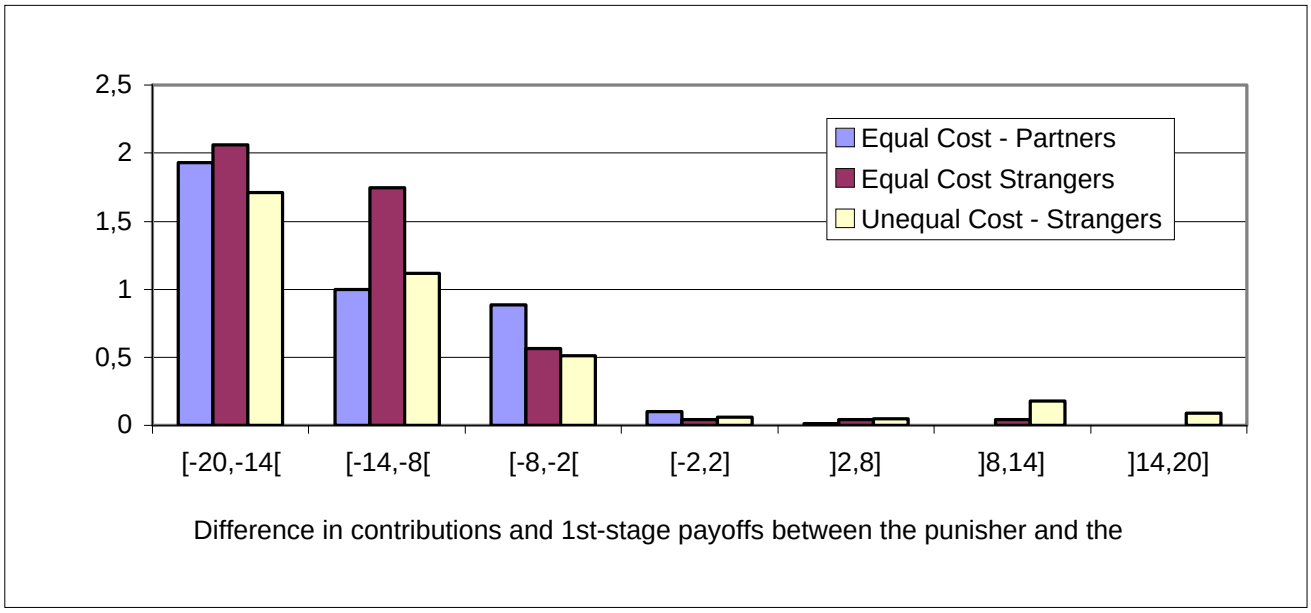

Fig.2. Punishment intensity as a function of payoff and contribution inequality between the punisher and the target (the percentages at the top of the bars represent the relative frequency of each category of deviation in the various treatments)

Appendix – Instructions of the Equal Cost Treatment [*other instructions are available upon request*]

You are now taking part in an economic experiment during which you can earn money. Your earnings depend on your decisions and the decisions of others. It is therefore very important that you read these instructions with care.

The instructions we have distributed to you are solely for your private information**. It is forbidden to communicate with the other participants during the experiment.** If you violate this rule, we shall have to exclude you from the experiment and from all payments.

During the experiment your entire earnings will be calculated in ECU (Experimental Currency Units). At the end of the experiment the total amount of ECU you have earned will be converted to Euros and paid in cash according to the following rules:

- Your final payoff in ECU consists of the sum of your payoffs in each period.
- This final payoff in ECU will be converted into Euros at the rate: 100 ECU = 2 Euros.
- In addition, you will be given a show up fee of 3 Euros.

The experiment is divided into successive periods. In each period the participants are divided into groups of four. You will therefore be in a group with 3 other participants. In each new period, the composition of your group will change: your probability of interacting more than one time with the same three other people is very low. You will not be informed of the identity of the other group members.

INSTRUCTIONS FOR THE FIRST 10 PERIODS

The four subjects belonging to a group can participate in a project, by contributing to a group account that will be shared among them. This amount of this group account is determined by the individual contributions of the four members of the group.

- At the beginning of each period each participant receives 20 ECU. In the following we will call this his or her endowment.
- Each of the four subjects simultaneously decides how much of his or her endowment he/she will contribute to the project. After having decided how much of the 20 ECU you want to contribute to the project, by choosing a number between 0 and 20, you must press the OK button Once you have done this, you can no longer change your decision for the current period.
- After all members of your group have made their decision your screen will show you the total amount of ECU contributed to the project by the four group members (including your own contribution). This screen shows you how many ECU you have earned for this period.
- Your income consist of two parts:
  - the amount of your endowment which you have kept for yourself (i.e. 20 – your contribution to the project),
  - the income from the project: this income represents 40% of the total contribution of all four group members to the project (the total includes your own contribution)

  Your income in ECU in each period is therefore:

(20-your contribution to the project) + 40%(total contributions to the project)

The income of each group member from the project is calculated in the same way, this means that each group member receives the same income from the project.

For example, suppose the total of the contributions of all group members is 60 ECU. In this case each member of the group receives an income from the project of 40% (60) = 24 ECU. If the total contribution to the project is 9 ECU, then each member of the group receives an income of 40% (9) = 3.6 ECU from the project.

For each ECU that you keep for yourself you earn an income of 1 ECU. For every ECU you contribute to the project instead, the total contribution rises by one ECU. Your income from the project will rise by 40% (1) = 0.4 ECU. The income of the other group members will also rise by 0.4 ECU each, so that the total income of the group from the project rises by 1.6 ECU. Your contribution to the project therefore also increases the income of the other group members. On the other hand you will earn money from each ECU contributed by other members to the project. For each ECU contributed by any member you earn 40% (1) = 0.4 ECU.

At the end of each period, groups randomly reshuffled.

If you have any question on these instructions, please raise your hand. We will answer your questions in private.

INSTRUCTIONS FOR THE FOLLOWING 10 PERIODS

[These instructions were distributed only at the end of the first 10 periods]

As before, groups are reshuffled randomly after each period. Each period consists of two stages.

In the first stage you have to decide how many ECU you would like to contribute to a project. In the second stage you are informed about the contributions of each of the three other group members to the project. You can decide whether or by how much to reduce their earnings from the first stage by distributing points to them. The following sections describe the activity in detail.

- **The first stage** is identical to the first ten periods.
  - At the beginning of each period each participant receives 20 ECU.
  - Each of the four subjects simultaneously decides how much of his or her endowment he/she will contribute to the project. After having decided how much of the 20 ECU you want to contribute to the project, by choosing a number between 0 and 20, you must press the OK button Once you have done this, you can no longer change your decision for the current period..
- At the end of this first stage, you will be individually informed of the amount contributed and of your income in the first stage which consists of two parts:
  - the amount of your endowment which you have kept for yourself (i.e. 20 – your contribution to the project),
  - the income from the project: this income represents 40% of the total contribution of all four group members to the project

  Your income of the first stage in each period is therefore:

(20-your contribution to the project)+ 40% (total contributions to the project)

The income of each group member from the project is calculated in the same way, this means that each group member receives the same income from the project.

- At the beginning of the **second stage**, your screen shows you how much each of your group members contributed to the project. You have now the possibility to reduce or leave unchanged the income of each group member by distributing points. You can distribute points to any member of your group to indicate your disapproval. Each point you distribute to a particular player lowers his or her payment according to the following schedule:

Number of points received from the same subject	0	1	2	3	4	5	6	7	8	9	10
Cost deduced from first-stage income of the subject who receives the points	0	1	2	4	6	9	12	16	20	25	30

Suppose for example that you give two points to one subject. This will reduce his first-stage income by 2 ECU; if you give 9 points to another subject, this will reduce his first-stage income by 25 ECU; if you give the last subject no points, this does not change his income. The amount of points you distribute to each subject determines therefore how much you reduce his income from the first-stage.

Similarly, your income can be modified if the other group members wish to do so.

- You are first informed how much each group member contributed to the project in the first stage. Please note that the order in which the contribution decisions of the three other subjects appear on your screen changes randomly each period (i.e. the first number which appears on your screen does not always correspond to the same subject).
- You must decide how many points to give to each of the other three group members to reduce their income or leave it unchanged. You must enter a value for each subject, between 0 and 10 points. If you do not wish to change the income of a specific subject, then you must enter 0.
- If you distribute points, you bear a cost in ECU which depends on the number of points you distribute to each subject. The more points you give to any subject, the higher your costs. Your total costs are equal to the sum of the costs of distributing points to each of the other three group members. The following table illustrates the relation between points distributed to a subject and the cost of doing so in ECU.

Number of points distributed to a subject	0	1	2	3	4	5	6	7	8	9	10
Cost deducted from your first-stage income for the distribution of points	0	1	2	4	6	9	12	16	20	25	30

Suppose for example that you give 2 points to one subject. This costs you 2 ECU. If you give 9 points to another subject, this costs you an additional 25 ECU; if you give the last subject no points, this has no cost for you. In this example, your total costs of distributing points would be 27 ECU (2 + 25 + 0). These costs will be displayed on your screen. As long as you have not pressed the OK button you can alter your decisions.

- Your final income in ECU in each period for the two stage is therefore calculated as follows:

= (income from the 1st stage) – cost of the points you have received - cost to you of points you distribute

Your income in ECU at the end of the second stage can be negative, if the costs of the points you distribute exceeds your income from the first stage. You can however avoid such losses with certainty through you own decisions.

* * *

If you have any question about these instructions, please raise your hand. We will answer your questions in private.

INSTRUCTIONS FOR THE LAST 10 PERIODS

[These instructions were distributed only at the end of the 20th period]

In the next ten periods, each period will follow the same rules as periods 1-10. The groups are reshuffled randomly after each period.

.

* * *

NOTES

[i] In the first stage of the game designed by Fehr and Gächter (2000), subjects contribute to a public good; in the second stage, after being informed about their individual contributions, the subjects can impose costly punishment on their team members. Contrary to the unique sub-game perfect Nash equilibrium of this game, subjects do punish their teammates whose level of contribution is lower than the average. The targets increase their contributions in reaction to punishment and the groups converge to the optimum of full cooperation.

[ii] The results of Fehr and Gächter (2000) on punishment have inspired many other studies that confirm the existence of peer pressure in groups and examine many variants of punishment. For example, Carpenter (2006) shows the influence of the group size on the efficiency of punishment. He also demonstrates that punishment is more efficient when agents are allowed to punish only a fraction of the group members. Bochet, Page, and Putterman (Forthcoming) show however that the efficiency of punishment on cooperation is lower than when communication is allowed. For their part, Fehr and Rockenbach (2003) identify a detrimental effect of sanctions; the crowding-out of norm-based motivations by punishment is also suggested by and Houser, Xiao, McCabe, and Smith (2005).

[iii] Fehr and Schmidt (1999) develop a model in which individuals exhibit inequality aversion if they dislike being better or worse than others. Bolton and Ockenfels (2000) propose another model of inequality aversion in which people compare their payoff to the average payoff of the group. Whereas the latter assumes that subjects like being as close as possible to the average payoff, Fehr and Schmidt assume that subjects dislike payoff differences compared to any other individual. Engelman and Strobel (2004) designed an experiment to compare the relative performance of these two theories and conclude in favor of Fehr and Schmidt.

[iv] Most papers which study the importance of inequality in public good games have focused on endowment heterogeneity. In contrast, our paper examines the relationship between punishment and inequality aversion when inequality arises endogenously from the subject's behavior.

[v] In our experiment, the cost to the target was higher than the cost to the punisher in all the cases where a subject punished another subject in the Unequal Cost Treatment. On average, the ratio is 2.74.

[vi] Alternatively, we could have imposed a fixed cost ratio between the punisher and the target higher than 1. We rejected this option in order to replicate the design in use in Fehr and Gächter (2000).

[vii] In all non-parametric statistical tests reported in this paper, the unit of observation is the group in the data collected under the partner matching protocol (N=6) and the session for the data collected under the stranger matching protocol (N=8).

[viii] The "negative deviation from punisher" variable is the absolute value of the difference between subject *j* 's contribution and the contribution of subject *i*. This variable is set equal to zero if the deviation is positive. The other deviation variables are constructed analogously.

[ix] The number of observations in which subjects punish individuals who contribute more than themselves is not sufficient to allow for a separate estimation of the model on this sub-sample.

[x] It should be noted that in standard linear public good games, the difference in contributions between two players is similar to the difference between their first-stage payoffs.

[xi] Here we consider the deviation from the average since a punished subject is only aware of the total number of points he received from the other group members and does not know who punished him.

[xii] In addition, when punishment is possible, we observe that average first-stage payoffs (i.e. before punishment decisions are made) are significantly higher than in treatments without punishment. The figures are 25.37 in the Unequal Cost Treatment, 24.22 in the Equal Cost Treatment under the stranger matching protocol and 27.83 under the partner matching protocol.